

2024

Разработчики:

ФГБОУ ВО «Гос.ИРЯ им. А.С. Пушкина», заведующий лабораторией когнитивных и лингвистических исследований, канд. фил. наук Лебедева Мария Юрьевна;

ФГБОУ ВО «Гос. ИРЯ им. А.С. Пушкина», научный сотрудник, канд. пед. наук Лапошина Антонина Николаевна

Протокол заседания Ученого Совета Гос. ИРЯ им. А.С. Пушкина № 53 от « 20 » декабря 20 24 г.

Ученый секретарь



Г.Н. Купцова

И.о. проректора-
начальника
управления
образовательной
деятельности



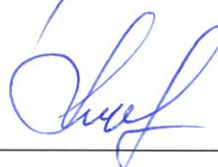
Е.Н. Чернышева

Проректор-начальник
управления
дополнительного
образования и организации
мероприятий



Д.А. Горбатова

И.О. начальника Центра
дополнительного
образования



М.А. Игнатьева

1 ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Направленность программы

Дополнительная общеразвивающая программа «Лингвистика для ИТ-отрасли: от лингвистических задач к большим языковым моделям» имеет техническую направленность.

Уровень программы

Уровень программы – вводный.

Актуальность программы обусловлена тем, что в условиях цифровизации общества умение работать с большими языковыми моделями является востребованным навыком. Программа помогает понять взаимосвязь между языком, языковыми структурами и современными технологиями. Программа знакомит учащихся с основами лингвистики, учит анализировать тексты и решать задачи, связанные с обработкой естественного языка. Это способствует развитию системного мышления и логики, а также готовит школьников к дальнейшему обучению в области информационных технологий. Кроме того, программа развивает навыки программирования и работы с данными, что делает её важной частью подготовки будущих специалистов в сфере ИТ.

Цель программы – формирование у школьников понимания принципов обработки естественного языка и применения этих знаний в практике работы с современными технологиями, базирующимися на больших языковых моделях, подготовка к осознанному выбору профессии в сфере информационных технологий.

Задачи программы

Обучающие:

- расширить и систематизировать научные знания о языке и базовых понятиях лингвистики; взаимосвязи его уровней и единиц;
- сформировать представление об автоматической обработке естественного языка, технологиях машинного обучения;
- сформировать и развить навыки эффективного поиска, формирования запросов к нейросети (промптоинга);
- дать представление о способах решения задач автоматической обработки естественного языка;
- сформировать представление о работе специалистов в области компьютерной лингвистики.

Развивающие:

- развить логическое мышление и навыки анализа;

- развить проектные умения, в том числе поисковые (исследовательские), рефлексивные, коммуникативные, презентационные умения;

- способствовать развитию устойчивого интереса к сфере лингвистики и информационных технологий.

Воспитательные:

- воспитывать коммуникативные навыки личности для совместной работы в коллективе, формировать общие этические нормы и правила поведения и межличностного общения, навыки эффективной и бесконфликтной работы;

- содействовать формированию общей культуры обучающихся;

- воспитывать ответственность, самостоятельность, усидчивость, трудолюбие, аккуратность, целеустремленность.

Учащиеся, для которых программы актуальна

Возраст обучающихся по данной программе: 14–18 лет и студенты учреждений среднего профессионального образования.

Формы и режим занятий

Очно-заочная (с применением полностью или частично ЭО и ДОТ).

Количество обучающихся в группе: до 25 человек.

Занятия проходят 3 раза в неделю по 2 академических часа.

Предусмотрен перерыв продолжительностью 10 минут в конце каждого учебного часа.

Срок реализации программы

Общее количество учебных часов, запланированных на весь период обучения – 72 академических часа, из них 54 ак.ч. самостоятельной работы.

Планируемые результаты

Предметные результаты

По итогам обучения по программе обучающиеся будут

знать:

- базовые понятия теоретической и компьютерной лингвистики;

- основные принципы автоматической обработки текстовых данных;

- принцип работы информационного поиска, основные принципы семантического поиска и его отличия от традиционного поиска;

- основы работы систем автоматической классификации текстов по заданным параметрам;

- принцип работы больших языковых моделей;

- основы работы с лингвистическими корпусами;

уметь:

- решать простые задачи автоматического анализа текстов;

- определять частотность слова в тексте и формировать список

ключевых слов;

- формулировать поисковые запросы на искусственном и естественном языках для получения релевантных результатов;

- осуществлять эффективный поиск в интернете и в Национальном корпусе русского языка,

- использовать диалоговые ассистенты для решения информационных задач

- формировать запросы к нейросети и оценивать эффективность выдачи.

Личностные результаты

(результаты воспитания и развития личности)

- осознание себя членом коллектива;

- осознанное выполнение правил поведения в различных образовательных ситуациях, публичных выступлениях, правил по технике безопасности;

- организованность, самостоятельность.

Метапредметные результаты

- умение самостоятельно планировать пути достижения целей и выбирать самые эффективные способы решения учебных и познавательных задач;

- умение создавать, применять и преобразовывать знаки и символы, модели и схемы для решения учебных и познавательных задач;

- умение определять понятия, создавать обобщения, устанавливать аналогии, классифицировать, самостоятельно выбирать основания и критерии для классификации, устанавливать причинно-следственные связи, строить логическое рассуждение, умозаключение (индуктивное, дедуктивное и по аналогии) и делать выводы;

- умение организовать сотрудничество и совместную деятельность с педагогом и сверстниками, работать индивидуально и в группе, находить общее решение и разрешать конфликты на основе согласования позиций и учета интересов;

- развитие компетентности в области использования информационно-коммуникационных технологий.

2 ФОРМЫ АТТЕСТАЦИИ И ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

Формы контроля

Реализация программы «Лингвистика для ИТ-отрасли: от лингвистических задач к большим языковым моделям» предусматривает входную диагностику, текущий контроль, промежуточную и итоговую аттестацию обучающихся.

Входная диагностика осуществляется в форме входного тестирования.

Текущий контроль включает следующие формы: практические задания, вопросы по теории.

Итоговая аттестация проводится в форме защиты проекта.

Основным механизмом выявления результатов воспитания является педагогическое наблюдение.

Публичная презентация образовательных результатов программы осуществляется в форме: презентации работы собственной программы.

Обучающимся, успешно освоившим программу и прошедшим аттестацию в форме, предусмотренной программой, выдается документ, подтверждающий освоение программы (в соответствии с локальными нормативными актами учреждения).

Средства контроля

Контроль освоения обучающимися программы осуществляется путем оценивания следующих параметров:

1. соответствие результатов освоения программы заявленным целям и планируемым результатам обучения;

2. соответствие процесса организации и осуществления программы установленным требованиям к порядку и условиям ее реализации;

Результативность обучения дифференцируется по трем уровням: низкий, средний, высокий.

При *низком уровне* освоения программы:

1. обучающийся не смог подготовить собственную программу / программа работает некорректно

2. рассказ обучающегося о цели проекта характеризуется серьезными недочетами.

При *среднем уровне* освоения программы обучающийся:

1. программа обучающегося характеризуется существенными недочетами;

2. рассказ обучающегося о цели проекта выполнен на удовлетворительном уровне.

При *высоком уровне* освоения программы обучающийся:

1. программа обучающегося работает корректно, с ее помощью можно получить запланированный результат;

2. рассказ обучающегося о цели проекта подготовлен на высоком уровне.

Позиции педагогического наблюдения:

- уровень овладения учебным материалом;

- коммуникативные компетенции и овладение навыком работы в коллективе.

3 СОДЕРЖАНИЕ ПРОГРАММЫ

Учебно-тематический план

№ п/п	Название раздела, темы	Количество часов				Формы аттестации (контроля) по разделам
		Всего	Теорети- ческих	Практи- ческих	Самостоятельная работа	

1.	Введение	4	1	1	2	Опрос, выполнение практических заданий
1.1.	Введение. Естественные и искусственные языки	1	1			
1.2.	Входная диагностика. Разбор заданий входной диагностики	1		1	2	
2.	Основы компьютерной лингвистики	48	4	6	38	Опрос, выполнение практических заданий
2.1	Основы автоматической обработки языка		3	3	24	
2.2	Поиск как задача компьютерной лингвистики		1	1	6	
2.3	Корпусная лингвистика. Практика работы с корпусами текстов.			2	8	
3.	Машинное обучение и лингвистика. Принципы работы нейросетей	18	1	3	14	Опрос, выполнение практических заданий
3.1	Машинное обучение и лингвистика. Принципы работы больших языковых моделей.		1		6	
3.2	Практика использования нейросетей. Основы промптинга.			3	8	
4.	Итоговое занятие	2		2		Презентация проектной работы
	Итого	72	6	12	54	

Содержание учебно-тематического плана

1 Раздел «Введение»

1.1 «Введение. Естественные и искусственные языки».

Теоретическая часть

Знакомство с обучающимися. Ознакомление с программой. Ключевые понятия лингвистики, язык как система. Понятие естественного и искусственного языка, примеры естественных и искусственных языков. (лекция 1 ак.ч.)

1.2 Входное тестирование. Разбор заданий входной диагностики

2 Раздел «Основы компьютерной лингвистики»

2.1 Основы автоматической обработки языка. Уровень графики. Уровень лексики. Уровень грамматики.

Теоретическая часть

Графематический анализ текста. Лингвистика и криптография. Алфавитное кодирование. Расстояние Левенштейна. Некоторые принципы работы автоматических корректоров текста. (лекция 1 ак. ч.).

Лексика и автоматическая обработка текста. Частотность слова. Ключевые слова текста. Основы векторной семантики. (лекция 1 ак. ч.).

Грамматика и автоматическая обработка текста. Грамматическая синонимия. Грамматическая омонимия и ее разрешение. (лекция 1 ак. ч.).

Практическая часть

Знакомство со средой программирования Python для решения задач по автоматической обработке текста. Практикум по автоматической обработке текста. Ввод текста, графематический, морфологический, семантический, синтаксический анализ в системах автоматического анализа текста (практические занятия 3 ак.ч.)

2.2 Поиск как задача компьютерной лингвистики

Теоретическая часть

История поиска в интернете. Принципы семантического поиска. Особенности формирования запроса на искусственном языке, переход к запросам на естественном языке (лекция 1 ак. ч.).

Практическая часть

Практикум по эффективному поиску и основам поисковой оптимизации (практическое занятие 1 ак.ч.).

2.3.Корпусная лингвистика. Практика работы с корпусами текстов

Практическая часть

Понятие лингвистического поиска. Национальный корпус русского языка, его возможности, компонента. Приемы поиска в корпусе, принципы

эффективного поиска. Ограничения корпуса как инструмента. (практические занятия 2 ак.ч.).

3 Раздел «Машинное обучение и лингвистика. Принципы работы нейросетей»

3.1 Машинное обучение и лингвистика. Принципы работы больших языковых моделей.

Теоретическая часть.

Знакомство с понятием машинного обучения. Задачи машинного обучения, решаемые в рамках компьютерной лингвистики. Большие языковые модели (лекция 1 ак. ч.).

3.2 Практика использования нейросетей. Основы промптинга

Обзор доступных нейросетей для работы с текстом. Промптинг: структура, проверка эффективности ответа, фактчекинг. (практические занятия 2 ак.ч.).

4 Итоговое занятие

Теоретическая часть. Подведение итогов.

Практическая часть. Презентация проекта.

4 ОРГАНИЗАЦИОННО-ПЕДАГОГИЧЕСКИЕ УСЛОВИЯ РЕАЛИЗАЦИИ ПРОГРАММЫ

Учебно-методические условия реализации программы

Реализация программы «Лингвистика для ИТ-отрасли: от лингвистических задач к большим языковым моделям» предполагает следующие формы организации образовательной деятельности: лекция, демонстрация, практическое занятие.

Программа может быть реализована с применением электронного обучения и дистанционных образовательных технологий с использованием систем дистанционного обучения.

При реализации программы используются следующие образовательные технологии: личностно-ориентированное обучение, технология сотрудничества, технологии развивающего обучения, игровые технологии.

Также могут быть использованы дистанционные образовательные технологии.

При реализации программы используются следующие методы обучения: словесные, наглядные, практические.

Воспитывающий компонент программы

История и традиции ФГБОУ ВО «Гос.ИРЯ им. А.С. Пушкина» в содержании образовательной деятельности.

Содержание воспитания:

- история ФГБОУ ВО «Гос.ИРЯ им. А.С. Пушкина», центра дополнительного образования и развитие информационно-компьютерных технологий в дополнительном образовании;

- уникальность и традиции и ФГБОУ ВО «Гос.ИРЯ им. А.С. Пушкина»;

- знаменитые выпускники ФГБОУ ВО «Гос.ИРЯ им. А.С. Пушкина»;

- героическая летопись ФГБОУ ВО «Гос.ИРЯ им. А.С. Пушкина».

Основы культуры по профилю деятельности в содержании образовательной деятельности.

Содержание воспитания:

- культура поведения в различных образовательных ситуациях

- культура публичного выступления, презентации, представления творческой, компьютерной работы;

- культура организации рабочего, компьютерного места;

- трудолюбие, работоспособность;

- бережное отношение к оборудованию и оснащению компьютерного кабинета;

Основы социальной культуры в содержании образовательной деятельности.

Содержание воспитания:

- общепринятые правила и нормы поведения в различных социальных ситуациях;

- правила взаимодействия в паре, группе, команде;

- способы поведения в спорных ситуациях;

- права и обязанности гражданина (с учетом воспитательных ситуаций);

- наиболее значимые исторические события, традиции страны, имена (в рамках ежегодных тем);

- гуманистические принципы в отношениях с окружающими;

- оптимизм.

Методы воспитания:

Методы формирования сознания личности (ум)	Рассказ, объяснение, разъяснение, беседа, лекция, пример, диспут, анализ ситуаций.
Методы организации жизнедеятельности и поведения воспитанников (воля)	Поручение, упражнение, приучение, создание воспитывающих ситуаций, педагогическое требование, инструктаж, иллюстрации и демонстрации
Методы стимулирования и мотивации деятельности и поведения (чувства)	Соревнование, поощрение, метод естественных последствий, дискуссия, эмоциональное воздействие.
Методы контроля и самоконтроля в воспитании	Педагогическое наблюдение, беседа, опросы, анализ результатов деятельности обучающихся, создание контрольных ситуаций.

Методы самовоспитания	Рефлексия, самоотчет, самоодобрение, самоосуждение.
-----------------------	---

Технологии воспитательной работы:

1. Проектная технология воспитания - вовлечение учащихся в совместные исследовательские проекты;
2. Технология цифрового гражданства - воспитании информационной грамотности и осознанного подхода к использованию технологий;
3. Технология ситуационного анализа - формирование ответственности и осознания социального влияния технологий.

Концептуальные идеи, принципы:

- коллективно – деятельностный подход к воспитанию: коллективное целеполагание, коллективная организация деятельности, коллективное творчество,
- эмоциональное насыщение жизни, организация соревновательности и игры в жизнедеятельности детей;
- личностный подход, одобрение социального роста детей.
- гуманно – личностная технология Ш. А. Амонашвили. Целевыми ориентациями являются:
 - способствование становлению, развитию и воспитанию в ребенке благородного человека путем раскрытия его личностных качеств;
 - развитие и становление познавательных сил ребенка;
 - идеал воспитания – самовоспитание.

Перечень методического обеспечения к программе

Для проведения занятий с применением электронного обучения и дистанционных образовательных технологий с использованием систем дистанционного обучения по каждой учебной теме разработаны информационные материалы и технологические карты (инструкции, памятки) по выполнению обучающимися практических заданий.

Материально-технические условия реализации программы

Наименование специализированных аудиторий, кабинетов, лабораторий	Вид занятий	Наименование оборудования, программного обеспечения
Аудитория	Лекции	компьютер, мультимедийный проектор, экран, доска
Компьютерный класс	Практические занятия	мультимедийный проектор, экран, персональный компьютеры на каждого обучающегося с доступом в интернет и предустановленным браузером на каждого обучающегося,

		программное обеспечение: Visual Studio Code (версия, наиболее подходящая с учетом требований операционной системы, установленной на ПК) с установленным Python 3.9 или выше, программное обеспечение: R (https://cran.r-project.org/bin/) и R Studio (https://www.rstudio.com/products/rstudio/download/).
--	--	---

Требования к помещению для занятий: просторное, светлое помещение.
Требования к мебели: парты/столы, стулья, соответствующие возрасту обучающихся.

Учебно-информационное обеспечение программы

Нормативно-правовые акты и документы

1. Федеральный закон от 29 декабря 2012 г. № 273-ФЗ "Об образовании в Российской Федерации" (с изм. на 24 июня 2023 года).
2. Концепция развития дополнительного образования детей до 2030 года (с изм. на 15.05.2023 г.) (утверждена распоряжением Правительства Российской Федерации от 31 марта 2022 г. № 678-р).
3. Порядок организации и осуществления образовательной деятельности по дополнительным общеобразовательным программам (утвержден приказом Министерства просвещения Российской Федерации от 27 июля 2022 г. № 629).
4. Целевая модель развития региональных систем дополнительного образования детей (утверждена приказом Министерства просвещения Российской Федерации от 3 сентября 2019 г. № 467) (с изм. на 21.04.2023).
5. Методические рекомендации по проектированию дополнительных общеразвивающих программ (включая разноуровневые программы): приложение к письму Министерства образования и науки Российской Федерации от 18 ноября 2015 г. № 09-3242.
6. Методические рекомендации по реализации дополнительных общеобразовательных программ с применением электронного обучения и дистанционных образовательных технологий: приложение к письму Министерства просвещения Российской Федерации от 31 января 2022 г. № ДГ-245/06.
7. СП 2.4.3648-20 «Санитарно-эпидемиологические требования к организации воспитания и обучения, отдыха и оздоровления детей и молодежи» (утверждены постановлением Главного государственного санитарного врача Российской Федерации от 28 сентября 2020 г. № 28).
8. СанПиН 1.2.3685-21 «Санитарные нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания» (утверждены постановлением Главного государственного санитарного врача Российской Федерации от 28 января 2021 г. № 2).
9. Приказ Департамента образования города Москвы от 17.12.2014 г. № 922 «О мерах по развитию дополнительного образования детей» (с изм. на 24.10.2022).

10. Приказ Департамента образования и науки города Москвы от 3.04.2023 г. № 271 «О внесении изменений в приказ Департамента образования и науки города Москвы от 17 декабря 2014 года № 922». - для программ вводного уровня.

Литература

1. Компьютерный анализ текста // Цифровые гуманитарные исследования: монография / А.Б. Антопольский, А.А. Бонч-Осмоловская, Л.И. Бородин и др. Красноярск: Сиб. федер. ун-т, 2023.

2. Автоматическая обработка текстов на естественном языке и анализ данных : учеб. пособие / Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лукашевич Н.В., Сапин А.С. — М.: Изд-во НИУ ВШЭ, 2017. — 269 с.

3. Прикладная и компьютерная лингвистика / Под ред. Николаева И.С. и др. — М.: ЛЕНАНД, 2016

4. Зализняк А.А. Грамматический словарь русского языка. — М., Русский язык, 1980.

5. Международные стандарты в области корпусной лингвистики. // Структурная и прикладная лингвистика. Выпуск 9. СПб., 2012 С. 201-221.

6. Ляшевская О.Н. и др. 2010. Оценка методов автоматического анализа текста: морфологические парсеры русского языка // Компьютерная лингвистика и интеллект. технологии: По материалам ежегодной Международной конференции «Диалог». Вып. 9 (16) — М.: Изд-во РГГУ.

7. Прикладная и компьютерная лингвистика / Под. ред. Николаева И.С., Митрениной О.В., Ландо Т.М. — М.: ЛЕНАНД, 2016. — 320 с.

8. Копотев М.В. Введение в корпусную лингвистику. Электронное учебное пособие для студентов филологических и лингвистических специальностей университетов. Praha: Animedia, 2014.

Интернет-ресурсы:

1. Портал «Образование на русском»;

2. URL: https://pushkininstitute.ru/external_courses/13;

3. Яндекс.Образование: Учебник по машинному обучению. URL: <https://education.yandex.ru/handbook/ml> ;

4. Национальный корпус русского языка. URL: <https://ruscorpora.ru/>;

5. Карта слов. URL: <https://kartaslov.ru/> ;

6. Использование расширенного синтаксиса запросов программным способом, URL: <https://learn.microsoft.com/ru-ru/windows/win32/search/-search-3x-advancedquerysyntax>;

7. Что такое поиск на естественном языке (NLS)? URL: <https://www.astera.com/ru/type/blog/what-is-natural-language-search-nls/>;

8. Переосмысление поиска: как новые диалоговые системы побеждают устаревшие LLM и традиционные поисковые системы без контекста, URL: <https://www.unite.ai/ru/redefining-search-how-emerging-conversational-engines-overcome-outdated-llms-and-context-less-traditional-search-engines/>;

Кадровое обеспечение программы

Программа «Лингвистика для ИТ-отрасли: от лингвистических задач к большим языковым моделям» реализуется квалифицированными научно-педагогическими кадрами Государственного института русского языка им. А.С. Пушкина, имеющим высшее педагогическое, филологическое и техническое образование, опыт преподавательской работы, в том числе в области, соответствующей профилю программы, на уровне основного и среднего общего образования, а также высшего образования, и постоянно повышающим уровень профессионального мастерства.